WHAT IS CLAIMED:

1. A method for screening molecular markers to identify classifiers, the method comprising:

a. obtaining, for members of a first training population, first molecular marker data reflective of the expression in blood of each of a plurality of molecular markers, wherein said first training population comprises a first trait subgroup and a second trait subgroup;

b. selecting a plurality of candidate molecular markers from among said plurality of molecular markers based on a determination of the ability of said first molecular marker data to discriminate between members of said first trait subgroup and members of said second trait subgroup;

c. obtaining, for members of a second training population, second molecular marker data reflective of the expression in blood of all or a portion of said plurality of candidate molecular markers, wherein said second training population comprises said first trait subgroup and said second trait subgroup;

d. generating a plurality of combinations of molecular markers from said candidate molecular markers;

e. generating a plurality of classifiers by applying a mathematical model to said second molecular marker data for each of said plurality of combinations of molecular markers; and

f. selecting one or more classifiers from said plurality of classifiers based on a determination of the ability of each classifier of said plurality of classifiers to discriminate between members of said first trait subgroup and members of said second trait subgroup.

2. The method of claim 1, wherein a subset of said plurality of candidate molecular markers of (b) is selected based on a determination of the ability of molecular marker data of said candidate molecular markers to discriminate between members of said first trait subgroup and members of said second trait subgroup.

3. The method of any of claims 1 to 2, wherein said determination of the ability to discriminate is made on the basis of a measure of statistical significance.

4. The method of any of claims 1 to 2, wherein said determination of the ability to discriminate is made on the basis of differential fold change.

5. The method of claim 3, wherein said determination of the ability to discriminate is further made on the basis of differential fold change.

6. The method of any of claims 4 or 5, wherein said selected candidate molecular markers demonstrate a differential fold change of greater than 2.0.

7. The method of any of claims 4 or 5, wherein said selected candidate molecular markers demonstrate a differential fold change of greater than 3.0.

8. The method of claim 3, wherein said determination of statistical significance is a p value and said p value is set such that the number of selected candidate molecular markers is less than 100.

9. The method of claim 3, wherein said determination of statistical significance is a p value and said p value is set such that the number of selected candidate molecular markers is less than 50.

10. The method of claim 3, wherein said determination of statistical significance is a p value and said molecular markers are selected if the molecular marker data results in a p value of less than 0.05.

11. The method of claim 3, wherein said determination of statistical significance is a p value and said molecular markers are selected the molecular marker data results in a p value of less than 0.01.

12. The method of any of claims 1 or 2, wherein said determination of the ability to discriminate is made on the basis of a Wald-Wolfowitz runs test, a Mann-Whitney U test, a Kolmogorov-Smirnov two-sample test, a Significant Analysis of Microarrays technique, or Manduchis' algorithm for assigning confidence to differentially expressed genes.

13. The method of any one of claims 2 to 12, wherein said determination of the ability of each of said subset of candidate molecular markers to discriminate between said members of said first trait subgroup and said second trait subgroup is made using second molecular marker data.

14. The method of claim 1, wherein said first training population and said second training population have zero or more members in common.

15. The method of claim 1, wherein said plurality of combinations includes all possible combinations of said candidate molecular markers.

16. The method of claim 1, wherein said plurality of combinations includes all possible combinations of two of said candidate molecular markers.

17. The method of claim 1, wherein said plurality of combinations includes all possible combinations of three of said candidate molecular markers.

18. The method of claim 1, wherein said plurality of combinations includes all possible combinations of four of said candidate molecular markers.

19. The method of claim 1, wherein said selecting one or more classifiers from said plurality of classifiers comprises:

i obtaining for each member of a scoring population, third molecular marker data reflective of the expression in blood of molecular markers within said plurality of classifiers, wherein said scoring population comprises members of said first trait subgroup and said second trait subgroup;

ii assigning a score, for each classifier in said plurality of classifiers, based on an ability of the respective classifier to discriminate between members of said first trait subgroup and members of said second trait subgroup in said scoring population using said third data; and

iii selecting one or more classifiers from among said plurality of classifiers based on the score assigned to the selected classifier.

20. The method of claim 19, wherein said scoring population and said first and second training populations have zero or more members in common and said third data corresponds with said first and second data accordingly.

21. The method of claim 19, wherein said selecting one or more classifiers based on score comprises:

i ranking each classifier in the plurality of classifiers on the basis of the score assigned to said classifier; and

ii selecting the top 10 ranking classifiers.

22. The method of any of claims 19 and 21, wherein said score, for each respective classifier in said plurality of classifiers, is a receiver operator curve (ROC) score determined by an area under a receiver operator curve obtained by applying the respective classifier to said scoring population.

23. The method of claim 22, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.5.

24. The method of claim 22, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.65.

25. The method of claim 22, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.80.

26. The method of claim 22, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.90.

27. The method of any of claims 1 to 26, wherein said first and second molecular marker data is obtained from one or more available databases.

28. The method of claim 1, wherein said first molecular marker data is obtained using a data collection method that allows for the collection of expression data corresponding to molecular markers for a major portion of the genome.

29. The method of claim 1, wherein said first molecular marker data is obtained using a microarray.

30. The method of claim 1, wherein said second molecular marker data is obtained using quantitative RT-PCR.

31. The method of claim 1, wherein said mathematical model is one of a regression model, a neural network, a clustering model, principal component analysis, nearest neighbor classifier analysis, linear discrimination analysis, quadratic discriminant analysis, a support vector machine, a decision tree, a genetic algorithm, a projection pursuit, or weighted voting.

32. The method of claim 31, wherein said mathematical model is optimized using bagging, boosting, or the Random Subspace Method.

33. The method of claim 1, wherein the number of candidate molecular markers selected comprises less than 100 molecular markers.

34. The method of claim 1, wherein the number of candidate molecular markers selected comprises less than 50 molecular markers.

35. The method of claim 1, the method further comprising:

a. obtaining, for a test subject, fourth molecular marker data reflective of the expression in blood of candidate molecular markers of said one or more selected classifiers;

b. applying said one or more classifiers to said fourth molecular marker data to thereby classify said test subject into either said first trait subgroup or said second trait subgroup.

36. The method of claim 35, wherein said fourth molecular marker data is received over the Internet from a remote source.

37. A method for identifying classifiers for a trait, the method comprising:

a. obtaining, for members of a training population, molecular marker data reflective of the expression in blood of all or a portion of a plurality of candidate

molecular markers, wherein said plurality of candidate molecular markers are those molecular markers in a Table selected from Tables 1A-7I; wherein said Table is selected based on said trait and said training population comprises at least a first trait subgroup and a second trait subgroup for said trait as disclosed in Table F.

b. generating a plurality of combinations of molecular markers from said candidate molecular markers;

c. generating a plurality of classifiers by applying a mathematical model to said molecular marker data for each of said plurality of combinations; and

d. selecting one or more classifiers from said plurality of classifiers based on a determination of the ability of said one or more classifiers to discriminate between members of said first trait subgroup and members of said second trait subgroup.

38. The method of claim 37 wherein a subset of said plurality of candidate molecular markers of the Table selected in (a) is selected based on a determination of the ability of the molecular marker data of said subset of candidate molecular markers to discriminate between members of said first trait subgroup and members of said second trait subgroup.

39. The method of claim 38 wherein said molecular marker data is said second data.

40. The method of claim 37, wherein said plurality of combinations includes all possible combinations of molecular markers wherein said molecular markers are those identified in the selected Table.

41. The method of claim 37, wherein said plurality of combinations includes all possible combinations of pairs of molecular markers wherein said molecular markers are those identified in the selected Table.

42. The method of claim 37, wherein said plurality of combinations includes all possible combinations of three molecular markers wherein said molecular markers are those identified in the selected Table.

43. The method of claim 37, wherein said plurality of combinations includes all possible combinations of four molecular markers wherein said molecular markers are those identified in the selected Table.

44. The method of claim 38, wherein said plurality of combinations includes all possible combinations of said subset of molecular markers of said selected Table of molecular markers.

45. The method of claim 38, wherein said plurality of combinations includes all possible combinations of pairs molecular markers of said subset of molecular markers.

46. The method of claim 38, wherein said plurality of combinations includes all possible combinations of three molecular markers of said subset of molecular markers.

47. The method of claim 38, wherein said plurality of combinations includes all possible combinations of four molecular markers of said subset of molecular markers.

48. The method of any of claims 37 or 38, wherein said determination of the ability to discriminate is made on the basis of a measure of statistical significance.

49. The method of any of claims 37 or 38, wherein said determination of the ability to discriminate is made on the basis of differential fold change.

50. The method of claim 48, wherein said determination of the ability to discriminate is further made on the basis of differential fold change.

51. The method of any of claims 49 or 50, wherein said selected candidate molecular markers have molecular marker data which demonstrate a differential fold change of greater than 2.0.

52. The method of any of claims 49 or 50, wherein said selected candidate molecular markers have molecular marker data which demonstrate a differential fold change of greater than 3.0.

53. The method of claim 48, wherein said determination of statistical significance is a p value and said p value is set such that the number of selected candidate molecular markers is less than 100.

54. The method of claim 48, wherein said determination of statistical significance is a p value and said p value is set such that the number of selected candidate molecular markers is less than 50.

55. The method of claim 48, wherein said determination of statistical significance is a p value and said molecular markers are selected if they have a p value of less than 0.05.

56. The method of claim 48, wherein said determination of statistical significance is a p value and said molecular markers are selected if they have a p value of less than 0.01.

57. The method of any of claims 37 or 38, wherein said determination of the ability to discriminate is made on the basis of a Wald-Wolfowitz runs test, a Mann-

Whitney U test, a Kolmogorov-Smirnov two-sample test, a Significant Analysis of Microarrays technique, or Manduchis' algorithm for assigning confidence to differentially expressed genes.

58. The method of any one of claims 38 to 57, wherein said determination of the ability of each of said subset of candidate molecular markers to discriminate between said members of said first trait subgroup and said second trait subgroup is made using said second data.

59. The method of claim 37, wherein said first training population and said second training population have zero or more members in common.

60. The method of claim 37, wherein said selecting one or more classifiers from said plurality of classifiers comprises:

i obtaining for each member of a scoring population, third data reflective of the expression in blood of molecular markers within said plurality of classifiers, wherein said scoring population comprises members of said first trait subgroup and said second trait subgroup;

ii assigning a score, for each classifier in said plurality of classifiers, based on an ability of the respective classifier to discriminate between members of said first trait subgroup and members of said second trait subgroup in said scoring population using said third data; and

iii selecting one or more classifiers from among said plurality of classifiers based on the score assigned to the selected classifier.

61. The method of claim 60, wherein said scoring population and said first and second training populations have zero or more members in common and said third data corresponds with said first and second data accordingly.

62. The method of claim 60, wherein said selecting one or more classifiers based on score comprises:

i ranking each classifier in the plurality of classifiers on the basis of the score assigned to said classifier; and

ii selecting the top 10 ranking classifiers.

63. The method of any of claims 60 and 62, wherein said score, for each respective classifier in said plurality of classifiers, is a receiver operator curve (ROC) score determined by an area under a receiver operator curve obtained by applying the respective classifier to said scoring population.

64. The method of claim 63, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.5.

65. The method of claim 63, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.65.

66. The method of claim 63, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.80.

67. The method of claim 63, wherein said selecting based on score comprises selecting those classifiers in said plurality of classifiers that have an ROC score greater than 0.90

68. The method of any of claims 37 to 67, wherein said molecular marker data is obtained from one or more available databases

69. The method of claim 37, wherein said molecular marker data is obtained using quantitative RT-PCR

70. The method of claim 37, wherein said mathematical model is one of a regression model, a neural network, a clustering model, principal component analysis, nearest neighbor classifier analysis, linear discrimination analysis, quadratic discriminant analysis, a support vector machine, a decision tree, a genetic algorithm, a projection pursuit, or weighted voting.

71. The method of claim 71, wherein said mathematical model is optimized using bagging, boosting, or the Random Subspace Method.

72. The method of claim 37, wherein the number of candidate molecular markers selected comprises less than 100 molecular markers.

73. The method of claim 37, wherein the number of candidate molecular markers selected comprises less than 50 molecular markers.

74. The method of claim 37, the method further comprising:

a. obtaining, for a test subject, second molecular marker data reflective of the expression in blood of candidate molecular markers of said one or more selected classifiers;

b. applying said one or more classifiers to said second molecular marker data to thereby classify said test subject into either said first trait subgroup or said second trait subgroup.

75. The method of claim75, wherein said second molecular marker data is received over the Internet from a remote source.

76. A system for analysing the blood of a test subject, the system comprising:

a. obtaining, for said test subject, data reflective of the expression in blood of each molecular marker related to a classifier generated according to the method of any of claims 1 or 37;

b. applying said classifier to said data to thereby classify said test subject into a first trait subgroup or a second trait subgroup.

77. A composition useful for diagnosing a trait of interest said composition comprising a plurality of isolated polynucleotides each of said plurality of isolated polynucleotides selectively hybridizing to a molecular marker product so as to permit said plurality of isolated polynucleotides to generate molecular marker data for a combination of molecular markers, wherein said combination of molecular markers are selected from a Table chosen from one of Tables 1A to 7I and wherein said combination of molecular markers are derived using the method of claim 63 and results in a ROC score of greater than 0.6.

78. The composition of claim 78 wherein said trait of interest is selected from those traits disclosed in Table F.

79. A system for analysing the blood of a test subject, the system comprising:

a. a biochemical device for obtaining, for a test subject, data reflective of the expression in blood of each molecular marker in a classifier derived according to the method of claim 1;

b. a computing device for applying said classifier to said data to thereby classify said test subject into either a first trait subgroup or a second trait subgroup; and

c. a display for indicating to a user the result of said classification.

80. A system for screening molecular markers to identify classifiers, the system comprising a processor and being characterized by:

a. means for obtaining, for members of a first training population, first data reflective of the expression in blood of each of a plurality of molecular markers, wherein said first training population comprises a first trait subgroup and a second trait subgroup;

b. means for selecting a plurality of candidate molecular markers from among said plurality of molecular markers based on a determination of the ability of said

molecular markers to discriminate between members of said first trait subgroup and members of said second trait subgroup using said first data;

c. means for obtaining, for members of a second training population, second data reflective of the expression in blood of all or a portion of said plurality of candidate molecular markers, wherein said second training population comprises said first trait subgroup and said second trait subgroup

d. means for generating a plurality of combinations of molecular markers from said candidate molecular markers;

e. means for generating a plurality of classifiers by applying a mathematical model to each of said plurality of combinations of molectular markers using said second data; and

f. means for selecting one or more classifiers from said plurality of classifiers based on a determination of the ability of each classifier of said plurality of classifiers to discriminate between members of said first trait subgroup and members of said second trait subgroup.

81. A system for identifying classifiers for a trait, the system comprising a processor and being characterized by:

a. means for obtaining, for members of a training population, data reflective of the expression in blood of all or a portion of a plurality of candidate molecular markers, wherein said plurality of candidate molecular markers are those molecular markers in a Table selected from Tables 1A-7I; wherein said Table is selected based on said trait and said training population comprises at least a first trait subgroup and a second trait subgroup for said trait.

b. means for generating a plurality of combinations of molecular markers from said candidate molecular markers;

c. means for applying a mathematical model to each of said plurality of combinations, using said second data, to derive a plurality of classifiers;

d. means for selecting one or more classifiers from said plurality of classifiers based on a determination of the ability of said plurality of classifiers to discriminate between members of said first trait subgroup and members of said second trait subgroup.